Climatic Change (2014) 127:353–369
DOI 10.1007/s10584-014-1254-5

Non-stationary extreme value analysis in a changing climate

Linyin Cheng · Amir AghaKouchak · Eric Gilleland · Richard W Katz

Received: 30 December 2013 / Accepted: 14 September 2014 / Published online: 24 September 2014

Abstract This paper introduces a framework for estimating stationary and non-stationary return levels, return periods, and risks of climatic extremes using Bayesian inference. This framework is implemented in the Non-stationary Extreme Value Analysis (NEVA) software package, explicitly designed to facilitate analysis of extremes in the geosciences. In a Bayesian approach, NEVA estimates the extreme value parameters with a Differential Evolution Markov Chain (DE-MC) approach for global optimization over the parameter space. NEVA includes posterior probability intervals (uncertainty bounds) of estimated return levels through Bayesian inference, with its inherent advantages in uncertainty quantification. The software presents the results of non-stationary extreme value analysis using various exceedance probability methods. We evaluate both stationary and non-stationary components of the package for a case study consisting of annual temperature maxima for a gridded global temperature dataset. The results show that NEVA can reliably describe extremes and their return levels.

1 Introduction

The Intergovernmental Panel on Climate Change (IPCC) Special Report on Managing the Risks of Extreme Events and Disasters (Field et al. 2012) stressed that continuation of the observed Earth warming would change the frequency, severity and spatial pattern of climatic extremes. Recently, climatic extremes have been widely studied at a range of spatial and temporal scales (Jakob 2013; AghaKouchak et al. 2013; Diffenbaugh and Giorgi 2012; Kharin et al. 2007; Easterling et al. 2000). Climatic extremes, including heavy precipitation events and extreme hot days, have substantially increased in the past few decades (Alexander et al. 2006; Vose et al. 2005). A recent study shows that even concurrent extremes (e.g., warm-dry and

Electronic supplementary material The online version of this article (doi:10.1007/s10584-014-1254-5) contains supplementary material, which is available to authorized users.

L. Cheng · A. AghaKouchak (✉)
University of California Irvine, Irvine, CA 92697, USA
e-mail: amir.a@uci.edu

E. Gilleland · R. W. Katz
National Center for Atmospheric Research, Boulder, CO 80307, USA

warm-wet conditions) have increased significantly in the second half of the 20^{th} century (Hao et al. 2013).

Under the assumption of a stationary climate, the concepts of return level and return period provide critical information for design, decision-making, and assessing the impacts of rare weather and climatic events (Rosbjerg and Madsen 1998). For example, the return level with a T-year return period represents an event that has a $1/T$ chance of occurrence in any given year (Cooley et al. 2007). Infrastructure design concepts have long relied on stationary return levels, which assume no change to the frequency of extremes over time (Klein et al. 2009). However, the frequency of extremes has been changing and is likely to continue changing in the future (Milly et al. 2008; IPCC 2007). Therefore, concepts and models that can account for non-stationary analysis of climatic and hydrologic extremes are needed (e.g., Cooley 2013; Salas and Obeysekera 2013; Parey et al. 2010).

Katz et al. 2002 present non-stationarity in extremes in terms of changing quantiles (termed “effective return levels”), which vary as a function of time to keep the occurrence probability of an extremal event constant. Alternatively, Rootzén and Katz 2013 introduced the concept of Design Life Level to quantify the probability of exceeding a fixed threshold during the design life of a project. A recent study describes an R-package developed for analysis of extremes based on the concept of effective return levels (extRemes 2.0). Another available R-package (GEVcdn: Cannon 2011) supplies a framework for a conditional density estimation network, and can be used to perform non-stationary extreme value analysis. However, these packages do not provide any non-stationary generalization of the concepts of return period and return level frequently used in hydrology.

The concept of return period can also be extended to a non-stationary framework (e.g., Rootzén and Katz 2013; Salas and Obeysekera 2013). In this study, we introduce a framework for non-stationary extreme value analysis for practical and effective analysis of climate extremes under both stationary and non-stationary conditions using Bayesian inference. The methods presented are available through a software package called Non-Stationary Extreme Value Analysis (NEVA). Under the non-stationary assumption, NEVA provides three different methods for estimation of return levels: (a) standard return levels (commonly used in hydrologic design) in which the exceedance probability is constant for any given return period during the life of the design (hereafter, design exceedance probability); (b) constant thresholds with time varying exceedance probability; and (c) effective return levels. A unique feature of NEVA is that it offers the associated posterior probability intervals and uncertainty bounds for the return level estimates under non-stationarity. These features make NEVA a practical and attractive tool for users from across different fields, especially climatology and hydrology, to analyze extremes under both stationary and non-stationary assumptions.

2 Extremes in a Non-stationary climate: theory

Extreme Value Theory (EVT) provides a rigorous framework for analysis of climate extremes and their return levels (Katz et al. 2002; Coles 2001). Under a wide range of conditions, the distribution of the maxima or minima converges to one of the three limiting distributions: Gumbel, Fréchet, or Weibull (Katz et al. 2002; Leadbetter et al. 1983; Gumbel 1958). The combination of these three distributions into one family is referred to as the Generalized Extreme Value (GEV) distribution. A variety of studies apply the GEV to analyze extremes (Katz 2013; Towler et al. 2010; AghaKouchak and Nasrollahi 2010; Beniston et al. 2007; El Adlouni et al. 2007; Villarini et al. 2009; Kharin and Zwiers 2005; Zhang et al. 2001; Smith 2001; Gumbel 1942). This technique is often referred to as the block maxima approach (e.g., Coles 2001). Another form of the EVT is known as the peak-over-threshold (POT) approach, in which extreme values above a high threshold are analyzed using a generalized Pareto distribution (Coles 2001; Smith 1987). Both annual maxima and POT are widely applied in studying climatic extreme events (Villarini et al. 2011; Li et al. 2005; Davison and Smith 1990).

The cumulative distribution function of the GEV can be expressed as (Coles 2001):

$$\Psi(x) = \exp\left\{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{\frac{-1}{\xi}}\right\}, \left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right) > 0 \tag{1}$$

The GEV distribution is flexible for modeling different behavior of extremes with three distribution parameters $\theta=(\mu,\sigma,\xi)$: (1) the location parameter (μ) specifies the center of the distribution; (2) the scale parameter (σ) determines the size of deviations around the location parameter; and (3) the shape parameter (ξ) governs the tail behavior of the GEV distribution. The limiting case of $\xi\rightarrow 0$ gives the Gumbel distribution, $\xi < 0$ the Weibull distribution and $\xi > 0$ the Fréchet distribution.

The extreme value theory for stationary random sequences has been extensively studied (Papalexiou and Koutsoyiannis 2013). In this study, stationarity is defined as time invariance of extremes' properties (Leadbetter et al. 1983). For a non-stationary process, the parameters of the underlying distribution function are time-dependent (Renard et al. 2013; Gilleland and Katz 2011; Katz 2010; Cooley 2009) and hence, the properties of the distribution would vary with time (Meehl et al. 2000). In NEVA, the location parameter is assumed to be a linear function of time to account for non-stationarity (Eq. 2), while keeping the scale and shape parameters constant:

$$\mu(t) = \mu_1 t + \mu_0 \tag{2}$$

where t is the time (in years), and $\beta=(\mu_1,\mu_0,\sigma,\xi)$ are the parameters. Alternative models may be used, such as polynomial trends, step changes, trends on the scale or the shape parameter, etc. (Renard et al. 2013). The methodology presented in this study can be used with different types of trends in location parameter. In hydrology and climate literature, the linear or log-linear models are usually preferred when searching for trends in the occurrence of extreme events (Beguería et al. 2011). While NEVA allows non-stationary σ and ξ ($\sigma(t)=\sigma_1(t)+\sigma_0 \xi(t)=\xi_1(t)+\xi_0$), in this study, only non-stationarity with respect to μ is discussed. The primary reason is that modeling temporal changes in σ and ξ reliably requires long-term observations that are often not available for practical applications.

NEVA detects the presence of trends and non-stationarity in extremes in historical data using the Mann-Kendall trend test (Kendall 1976; Mann 1945) at the user's choice of significance level. The default significance level is α=0.05, which is widely used in hydrological research (Zhang et al. 2004). This nonparametric rank-based test avoids making an assumption about the underlying distribution function (e.g., assuming the data is normally distributed) of hydrological variables (Kundzewicz and Robson 2004). The null hypothesis of no trend is rejected if the test statistic $|Z_S|$ is larger than the critical value $Z_{\alpha/2}$. The test returns either 0 when $|Z_S| \leq Z_{\alpha/2}$ (the null hypothesis of no trend cannot be rejected) or 1 when $|Z_S| > Z_{\alpha/2}$ (the null hypothesis of no trend is rejected). If the null hypothesis is not rejected, NEVA will perform extreme value analysis under the stationary assumption. Upon detection of a trend at the 5 % significance level (α=0.05), the GEV distribution parameters will be estimated under the non-stationary assumption (Eq. 2). This will allow estimating return values in a more realistic way consistent with the behavior of climatic extremes.

NEVA uses a Bayesian technique to infer the GEV distribution parameters under stationary and non-stationary conditions. The Bayesian-based Markov chain Monte Carlo (MCMC) approach for obtaining the posterior distribution of parameters from an arbitrary distribution has become increasingly popular and used in several studies of extremes (Stephenson and

Tawn 2004; Coles and Powell 1996). This approach combines the knowledge brought by a prior distribution and the observation vector $\overrightarrow{y} = (y_t)_{t=1:N_t}$ (Eqs. 4 and 5) into the posterior distribution of parameters $\theta=(\mu,\sigma,\xi)$. Here, N_t denotes the number of observations (e.g., annual maxima) in the observation vector $\overrightarrow{y}$. The default priors for the location and scale parameters are non-informative normal distributions, whereas the default for the shape parameter is a normal distribution with a standard deviation of 0.3 as suggested in Renard et al. 2013. The default priors can be changed to informative priors, and other choices of distribution functions can be used in NEVA. The prior distributions for all parameters are assumed to be independent.

Assuming independence between observations, the Bayes theorem for estimation of GEV parameters under the non-stationary assumption can be expressed as (Renard et al. 2013, 2006; Coles 2001):

$$p\left(\beta\middle|\overrightarrow{y},x\right)\propto p\left(\overrightarrow{y}\middle|\beta,x\right)p\left(\beta\middle|x\right) \tag{3}$$

$$p\left(\overrightarrow{y}\middle|\beta,x\right) = \prod_{t=1}^{N_t} p\left(y_t\middle|\beta,x(t)\right) = \prod_{t=1}^{N_t} p\left(y_t\middle|\mu(t),\sigma,\xi\right) \tag{4}$$

where $\beta=(\mu_1,\mu_0,\sigma,\xi)$ are the parameters. The stationarity can be treated as a special case of the above equation without $x(t)$:

$$p\left(\theta\middle|\overrightarrow{y}\right)\propto p\left(\overrightarrow{y}\middle|\theta\right)p(\theta) = \prod_{t=1}^{N_t} p\left(y_t\middle|\theta\right)p(\theta) \tag{5}$$

where $x(t)$ denotes the set of all covariate values under the non-stationary assumption. The resulting posterior distributions $p(\theta|\overrightarrow{y})$ and $p(\beta|\overrightarrow{y},x)$ provide information about parameters under stationarity $\theta=(\mu,\sigma,\xi)$ or non-stationarity $\beta=(\mu_1,\mu_0,\sigma,\xi)$. The entire process for inferring distribution parameters in NEVA is summarized in Fig. 1 and Fig. 2 for stationary and non-stationary conditions, respectively. NEVA generates a large number of realizations from the parameter joint posterior distribution using the Differential Evolution Markov Chain (DE-MC) (Vrugt et al. 2009; Ter Braak and Vrugt 2008; Ter Braak 2006). The DE-MC utilizes the genetic algorithm Differential Evolution (DE) for global optimization over the parameter space with the MCMC approach. The DE-MC's simplicity, speed of calculation, and convergence make it favorable over the conventional MCMC (Ter Braak 2006).

The main motivation for combining DE-MC with Bayesian inference is that one can obtain the posterior probability intervals (uncertainty bounds) of estimated return levels taking into account the uncertainty in all model parameters (non-stationary: μ_0, μ_1, σ, ξ; stationary: μ, σ, ξ). It is worth noting that NEVA assesses convergence of the sampling approach statistically. A method known as the criterion $\widehat{R}$, suggested by (Gelman and K. Shirley 2011), is built into NEVA as a convergence check. This method suggests that the $\widehat{R}$ values should remain below the critical value of 1.1 (see Gelman and K. Shirley 2011 for more details on computing $\widehat{R}$).

In addition to the Mann-Kendall test, the likelihood-ratio test can be used to compare the fit of the two nested models: the *null* model is the stationary (no trend) case (L_{Null}), whereas the *alternative* is the non-stationary (linear trend) case ($L_{Alternative}$). The log-likelihood ratio can be expressed as (Coles 2001):

$$D = -2\ln\frac{L_{Null}}{L_{Alternative}} \tag{6}$$

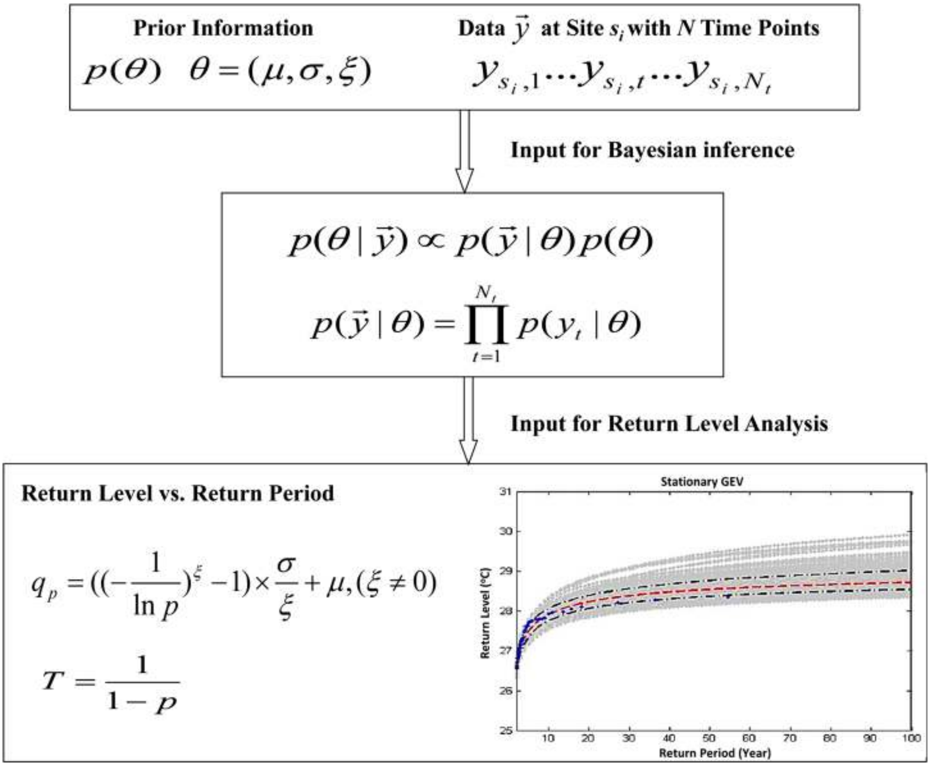

Fig. 1 NEVA's stationary GEV framework for extreme value analysis. The outputs are return levels versus return periods

In this paper, the modes of the posterior distributions are used as the parameter estimates for the likelihood ratio test statistic instead of the maximum likelihood estimates to ensure the test is still valid for Bayesian analysis. The likelihood ratio can then be used to test (e.g., at the $\alpha=0.05$ significance level) whether to reject the null model in favor of the alternative. The test returns either 0 when the non-stationary model does not fit significantly better than the stationary model or 1 when the non-stationary model fits significantly better than the stationary model. Note that the Mann-Kendall and likelihood ratio tests are both testing for trends, but under different assumptions: the Mann-Kendall test allows for non-linear trends in the location parameter and any form of distribution, while the likelihood ratio test assumes GEV distribution and only allows for a linear trend in the location parameter.

In order to further evaluate the fit of the *null* model M_1 (i.e., the stationary case), and the *alternative* model M_2 (i.e., the non-stationary case) based on the posterior distributions of sampled parameters, the Bayes factor is computed as:

$$K=\frac{\Pr\left(DA\middle|M_1\right)}{\Pr\left(DA\middle|M_2\right)}=\frac{\int\Pr\left(\theta_1\middle|M_1\right)\Pr\left(DA\middle|\theta_{1,}M_1\right)d\theta_1}{\int\Pr\left(\theta_2\middle|M_2\right)\Pr\left(DA\middle|\theta_{2,}M_2\right)d\theta_2} \tag{7}$$

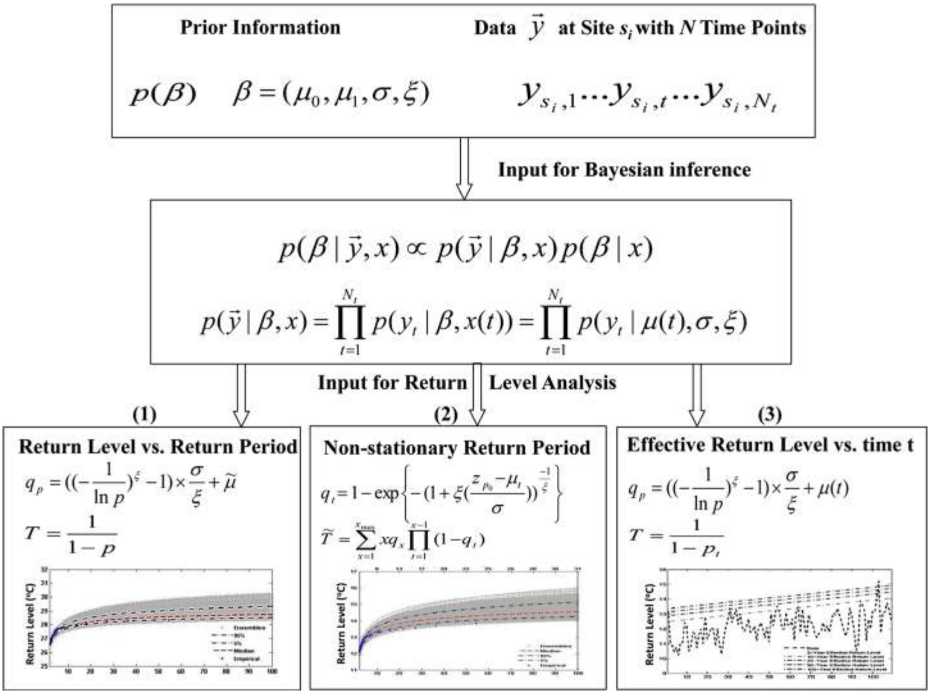

Fig. 2 NEVA's non-stationary GEV framework for extreme value analysis. The model outputs include: (1) standard return levels with design exceedance probability; (2) standard return levels with time varying exceedance probability; and (3) effective return levels

where DA denotes input data, and θ stands for model parameters. The term $\Pr(DA|M)$ can be expressed using a Monte Carlo integration estimation as $\Pr(DA|M) = \left\{\frac{1}{m}\sum_{i=1}^{m}\Pr\left(DA|\theta^{(i)},M\right)^{-1}\right\}^{-1}$, where m is the sample size (see Kass and Raftery 1995 for more details). A value of $K < 1$ indicates that the non-stationary model (M_2) fits better than the stationary model (M_1). Having multiple tests to detect stationarity or non-stationarity allows a more rigorous assessment of the goodness-of-fit.

While the original NEVA is designed for analysis of maxima in time series, users can apply NEVA for analysis of time series minima using the following transformation (Coles 2001):

$$\min(X_1,\ldots,X_n) = -\max(-X_1,\ldots,X_n) \tag{8}$$

where $X_1,\ldots,X_n$ is a time series of i.i.d. random variables.

Using the GEV distribution, NEVA computes the return periods and return levels of extremes (see Eqs. 9 and 10). In this approach, return levels are expressed as a function of the return period T (Cooley 2013):

$$T = \frac{1}{1-p} \tag{9}$$

where p is the non-exceedance probability of occurrence in a given year (assumed constant under stationarity). The $p-$ return level q_p derived from the GEV distribution can be expressed as (Coles 2001):

$$q_p = \left(\left(-\frac{1}{\ln p}\right)^{\xi} - 1\right) \times \frac{\sigma}{\xi} + \mu, \ (\xi \neq 0) \tag{10}$$

In NEVA, the time-variant parameter ($\mu(t)$) can be derived using different quantiles from the DE-MC. For example, in this paper, $\mu(t)$ is computed as: (a) median of $\mu(t)$ (refers to the effective return level for the year corresponding to the midpoint of the time series), and (b) 95 percentile of the DE-MC sampled $\mu(t)$ values. The latter can be considered a low risk (more conservative) approach for extreme value analysis by taking the 95 percentiles of the $\mu(t)$ values in historical observation to be used for future analysis (e.g., the effective return level for a year near the end of the record). The model parameters will then be used to estimate the non-stationary return levels as follows:

$$\widetilde{\mu} = Q_k(\mu_{t1}, \mu_{t2}, \ldots, \mu_{tn}), \ (\mu(t) = \mu_1 t + \mu_0) \tag{11}$$

$$q_p = \left(\left(-\frac{1}{\ln p}\right)^{\xi} - 1\right) \times \frac{\sigma}{\xi} + \widetilde{\mu}, \ (\xi \neq 0) \tag{12}$$

where κ=0.5 returns the median of n location parameters ($\mu_{t1}, \mu_{t2}, \ldots, \mu_{tn}$), and κ=0.95 corresponds to the 95 percentile of location parameters (a high quantile $\widetilde{\mu}$ indicating low risk extreme value analysis). This approach is similar to the stationary case, but allows for considering changes in the location parameter over time. This concept is termed design exceedance probability in this paper.

In a recent study, Salas and Obeysekera 2013, proposed another non-stationary counterpart of stationary return levels. In this approach, the probability that the first extreme event exceeding a given fixed threshold will occur at time x=1 is denoted by q_1, and the probability that it will occur at time x=2 is $(1-q_1)q_2$, and so forth (exceeding probabilities $q_1, q_2, q_3, \ldots, q_t$ vary over time). With the time varying exceedance probabilities q_t, a non-stationary concept determining the expected return period of the extreme event is outlined in Salas and Obeysekera 2013. This concept is based on the expected waiting time until the first exceedance of a fixed threshold, with the expected waiting time is calculated for time varying exceedance probabilities. In NEVA, the proposed DE-MC-Bayesian approach is integrated into Salas and Obeysekera 2013 to provide an alternative approach for non-stationary return level-return period analysis with time varying exceedance probability. The parameter estimation, uncertainty assessment, and sampling approach, as well as the log-likelihood test and Bayes factor computation remain similar in both design exceedance probability and time varying exceedance probability methods.

3 Results

In the following, NEVA is used for stationary and non-stationary extreme value analysis of annual temperature maxima from the Climatic Research Unit (CRU, New et al. 2000) gridded monthly temperature data (1901–2009). Fig. 3 displays the areas in which temperature block maxima exhibit a significant trend at the 5 % level and hence, non-stationary behavior (see

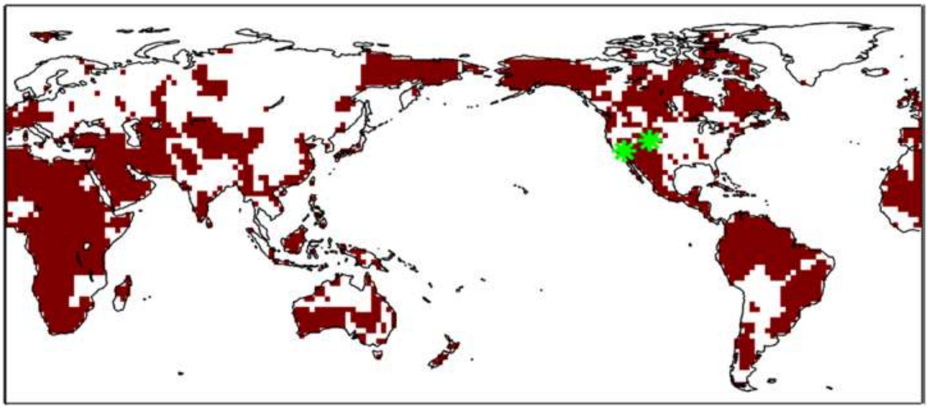

Fig. 3 Global Mann-Kendall Trend Analysis (Significant trend in red; No significant trend in white). The Star-marked locations are the pixels selected for time series analysis

dark red pixels in Fig. 3). The white land areas correspond to locations that do not show a significant trend in the annual temperature maxima. NEVA utilizes the suggested non-stationary extreme value analysis algorithm (Fig. 2) for the dark red pixels in Fig. 3, and the stationary algorithm (Fig. 1) in the rest of the pixels. The appropriate type of GEV (stationary or non-stationary) is fitted to each grid of monthly temperature maxima separately (i.e., not taking into account any spatial dependence or similarity of the trend at adjacent grid points).

Figure 4 shows the global annual temperature maxima return levels for the 5-year (4a) and 100-year (4b) return periods. As mentioned earlier, NEVA generates an ensemble of estimates based on DE-MC sampling. The median of the ensemble is used as the final return level values shown in Fig. 4. The uncertainty bounds of the computed return levels can be derived based on 5 % and 95 % posterior probability intervals of the ensemble as discussed below.

To further explore NEVA's outputs, two pixels in the central (Latitude 40.02° N, Longitude 105.27° W) and western (Latitude 34.05 ° N, Longitude 118.24 ° W) United States are selected for more detailed analysis (see green stars in Fig. 3). The two locations are close to urban areas in Boulder, CO and Los Angeles, CA where long-term observation stations have been available. In both locations, the Mann-Kendall trend test confirms the presence of non-stationarity at the 5 % significance level (see Figure S1 in Supplementary Material). The initial goodness-of-fit of the GEV model is assessed using Quantile-Quantile (Q–Q) plots of fitted and observed temperature maxima (see Figure S2 in Supplementary Material). The plot of the return levels versus the corresponding return periods at the two selected locations under both stationary (ignoring the observed trend) and non-stationary assumptions are displayed in Fig. 5 and Fig. 6. In both figures, the top panels (a) show return levels under the stationary assumption, while panels (b) exhibit non-stationary return levels for the observation period (here, 1901–2009). Panels (c) and (d) display non-stationary return levels for 100 years beyond observations (e.g., 2010–2109) using median and 95 percentile of sampled location parameters, respectively (see Eqs. 11 and 12). Consequently, panels (d) in Figs. 5 and 6 are more conservative estimates of future extreme return levels, and are termed as low risk (hereafter LR).

In the central U.S. (Fig. 5a), under the stationary assumption the posterior probability bounds do not encompass the empirical return levels, which indicates the assumptions for this model are not met. On the other hand, for the non-stationary model (Fig. 5b), the posterior probability bounds enclose the empirical return levels, indicating reasonable simulations (see the zoom in Fig. 5b). The selected point in the western U.S. (Fig. 6) exhibits a similar behavior.

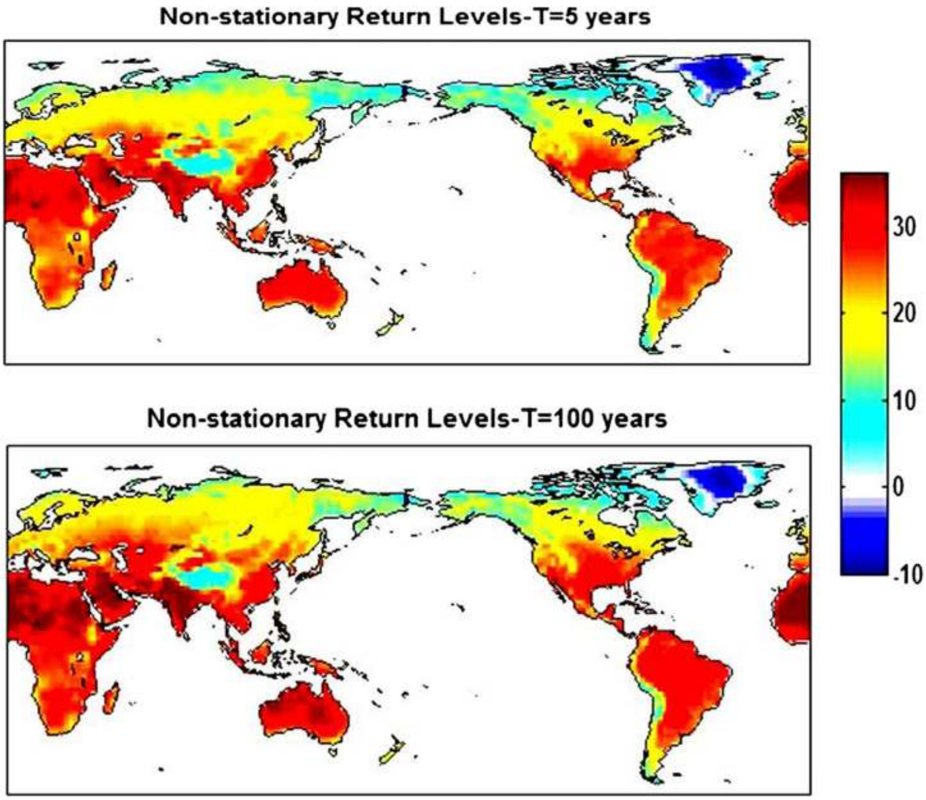

Fig. 4 5-year **a** and 100-year **b** annual monthly temperature maxima return levels (°C) under the non-stationary assumption, derived using the standard return levels with k=0.5 in Eq. 11

The non-stationary envelope of simulations (Fig. 6b) encompasses all of the empirical return levels, while the stationary ensemble (Fig. 6a) does not enclose all the points, including few points at the beginning and the last observation.

In the central U.S., the return levels under the stationary assumption (Fig. 5a) are much lower than those under the non-stationary assumption (Figs. 5c, d). For example, the return levels corresponding to the 50-year annual temperature maxima (ensemble median - red dashed lines) are 14.3° C, 15.5° C and 16.1° C, under stationary, non-stationary, and LR non-stationary, respectively. This result indicates that an unrepresentative assumption of stationarity would lead to misinterpretation (in this example, underestimation) of extreme climatic conditions. Another example is the pixel in the western U.S., where the positive trend is not as strong as the one in the central U.S. (compare Figures S1a and S1b). Nonetheless, if the observed linear trend continues in the future, the return levels will be underestimated under the stationary assumption. Considering a 50-year return level (ensemble median - red dashed lines), it is 28.5° C (stationary), 29.1° C (non-stationary), and 29.4° C (LR non-stationary). It should be noted, that the annual maxima is based on mean monthly temperature values and the daily maxima may exceed these values.

Once the parameters are sampled and return levels are simulated, the non-stationarity assumption included in the location parameter is tested using the log-likelihood and Bayes

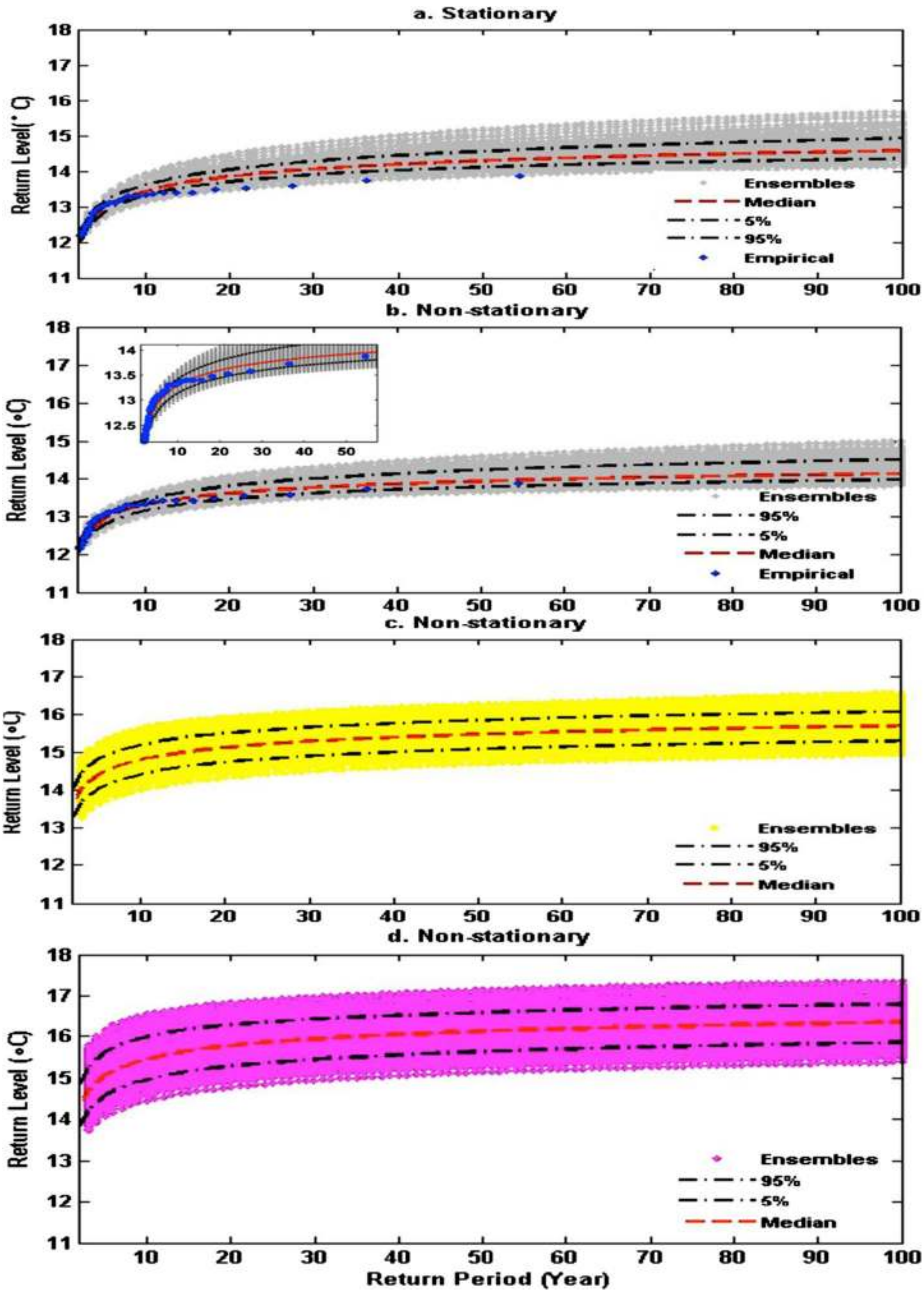

Fig. 5 Annual monthly temperature maxima return level vs. return period in the selected point in the central U.S. under stationary **a**, non-stationary during the period of observations 1901–2009 **b**, non-stationary based on median of sampled parameters **c**, and non-stationary based on the 95 percentile of the sampled parameters or Low Risk (LR) non-stationary **d**

factor approach discussed in Eq. 6 and 7. In both the central and western U.S., the log-likelihood test and Bayes factor confirm that the simulations exhibit non-stationary behavior

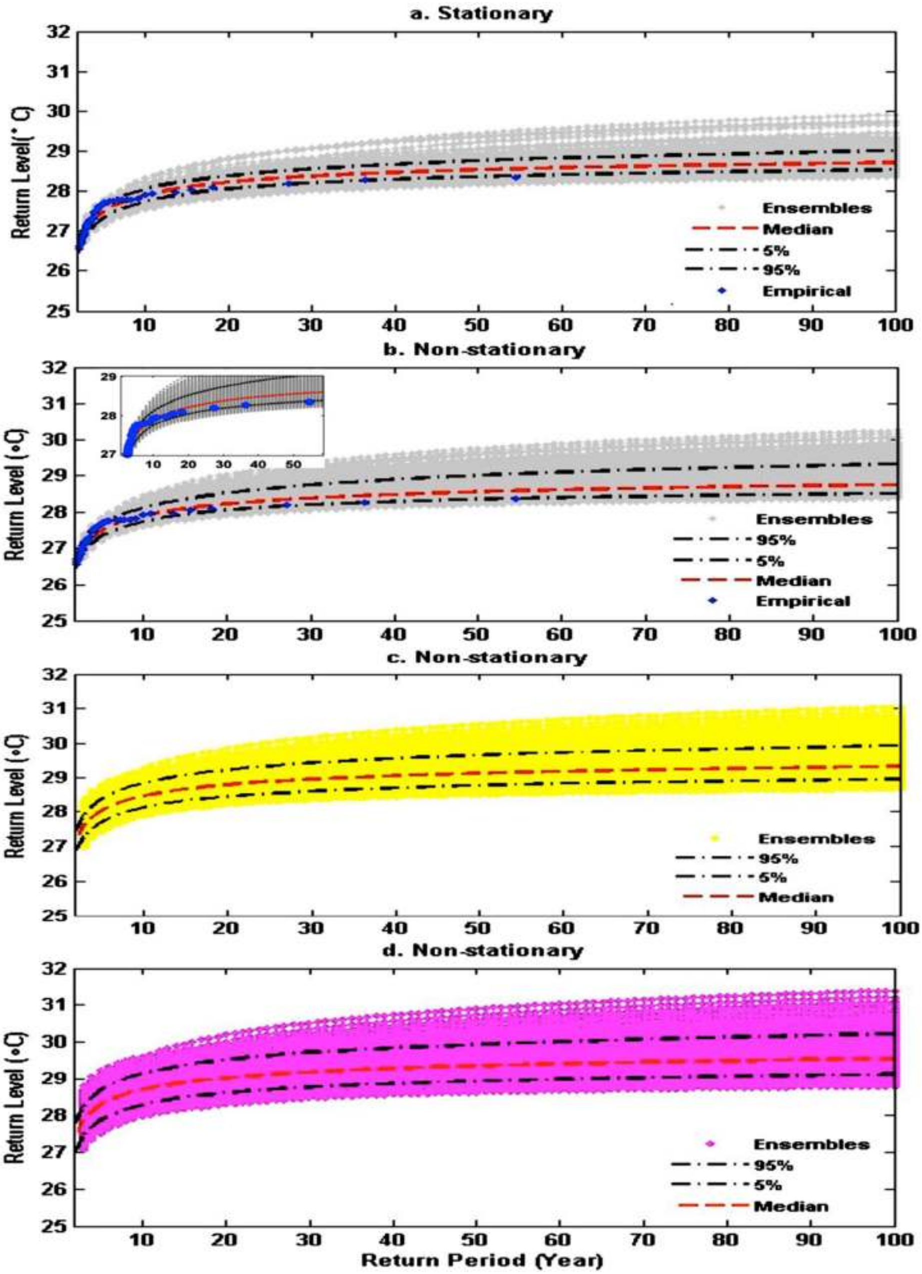

Fig. 6 Annual monthly temperature maxima return level vs. return period in the selected point in the western U.S. under stationary **a** non-stationary during the period of observations 1901–2009 **b**, non-stationary based on median of sampled parameters **c**, and non-stationary based on the 95 percentile of the sampled parameters or Low Risk (LR) non-stationary **d**

consistent with the Mann-Kendall test results (p-values smaller than the 0.05 significance level)—See Table S2 in Supplementary Materials.

As shown in the model flowchart (Fig. 2), NEVA can provide non-stationary return periods based on a time varying exceedance probability. Figure 7 presents return period vs. return level under stationarity and the corresponding non-stationary return periods for the two selected points in the central and western U.S. In this framework, the exceedance probability q_t varies through time. Since temperature extremes exhibit an upward trend at both locations, the exceedance probability q_t will increase over time. The probability distribution of the waiting time for the first extreme event to exceed a given threshold is a generalization of the geometric distribution, which enables determining the expected return period. For instance, in the western U.S., the 50-year return period under the stationary assumption corresponds to an approximate 30-year return period under a non-stationary condition (see Fig. 7b). In other words, an

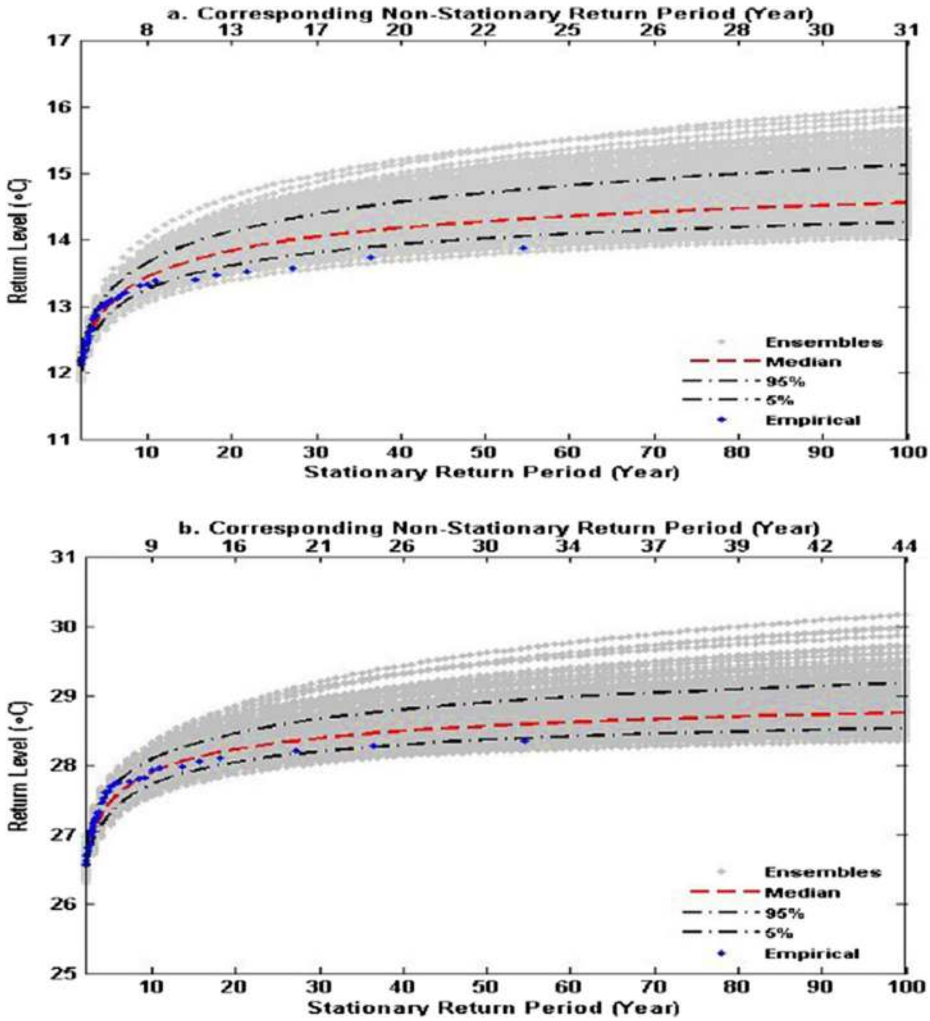

Fig. 7 Annual monthly temperature maxima return levels vs. return period under stationary (*bottom axes*) and the corresponding non-stationary (*top axes*) assumption at the selected points in the central **a** and western **b** United States

exceedance probability of 0.02 will increase to around 0.03 in a non-stationary climate. Similarly, in the central U.S., a 50-year extreme changes to a 22-year event in a changing climate. This framework allows displaying stationary and non-stationary return levels against each other.

As shown in the model flowchart (Fig. 2), NEVA generates non-stationary return levels based on both the standard definition in hydrology and the concept of effective return level. As an example, Fig. 8 demonstrates effective return levels for the two selected points in the central and western U.S. The figures show return levels versus the time covariate used in the linear regression (Eq. 2). In this concept, the return levels vary over time such that the probability of occurrence remains constant. Basically, the effective return level indicates what return level should be used for all years to have the same risk. In the western U.S., the effective return level corresponding to a 50-year (0.02 probability of occurrence) event during 1901–1950 is 28.5◦ C; whereas the same risk for a 100-year period (1901–2000) would be 28.8◦ C. Similarly, for another 100-year period (e.g., 2001–2100) the 50-year event would be different (here, 29.3◦ C). By providing both the standard and the effective return levels, as well as the integrated time-varying exceedance probability non-stationary return periods, NEVA allows the users to use the one that fits their application.

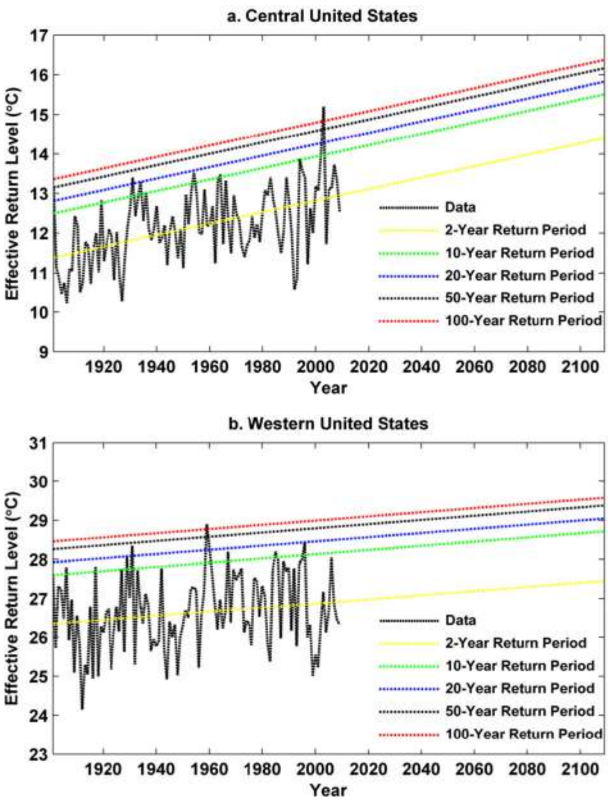

Fig. 8 Effective return levels under the non-stationary assumption at the selected points in the central **a** and western **b** United States

4 Conclusions

Substantial evidence shows that the climate is non-stationary, possibly due to anthropogenic climate change. The assumption of stationarity in extreme value analysis is therefore questionable and statistical models that explicitly allow for non-stationarity are much needed. Specifically, statistical models that can provide estimates of return levels under non-stationary conditions are essential for design and risk assessment purposes. In this study, a practical package named Non-stationary Extreme Value Analysis (NEVA) package is introduced for assessing extremes in a changing climate.

NEVA offers a framework for estimating non-stationary return levels, return periods, and risks of climatic extremes using Bayesian inference. In this approach, the model parameters are estimated using a Differential Evolution Markov Chain (DE-MC) for global optimization over the real parameter space with the Markov Chain Monte Carlo (MCMC) approach. NEVA also provides the posterior probability intervals (uncertainty bounds) of estimated return levels by combining DE-MC with Bayesian inference. A unique feature of the model is non-stationary extreme value analysis using both design exceedance probability and time varying exceedance probability methods.

The features and capabilities of NEVA can be summarized as follows: (a) the framework assesses trends in the observations; (b) depending on the trend, it performs stationary or non-stationary analysis of extremes and can test which model describes the data more appropriately based on the model outputs; (c) it provides non-stationary return levels based on three methods including one that resembles the standard approach in hydrology under stationarity, one based on expected waiting time with time varying exceedance probability, and effective return levels; and (d) NEVA includes a sampling framework that leads to uncertainty bounds of simulations. The return level and return period estimates can be used in hydrology and climate studies to assess the risk (probability of occurrence) of extremes.

By providing posterior probability intervals (e.g., 5 % and 95 % quantiles), NEVA offers a range of return levels, and the user can select the upper bound (low risk) or the lower bound (high risk) depending on the application at hand. Users can change the upper and lower bound quantiles of the simulated ensembles and also the significance level of the trend analysis component. Both stationary and non-stationary components of the package are evaluated using Climatic Research Unit (CRU) observations. The results indicate that NEVA simulates GEV-based return levels consistent with empirical observations. While the focus of this paper is on climate extreme value analysis, the methodology can potentially be used in different areas (hydrology, ecology, and economics) and with different data sets.

The authors stress that extreme value analysis and trend detection requires long-term and reliable data. One could detect a significant trend based on, say, 10 years of data taken from start and end points of opposite phases of an El Niño event. However, such a trend may not persist in the future. In fact, there is no guarantee that future trends will be similar to those estimated from the short-term or long-term past data. For time-varying probability of exceedances to be accepted as a simple extrapolation of what results from historical trends, one has to rely on the inertia of the climate system (i.e., typically few decades). To understand when the inertia of the system stops, one should rely on other tools such as dynamical model simulations. Therefore, care should be taken in extrapolating historical trends into the future, especially for long-term projections of extremes. On the other hand, long term projections may be affected by very different natural and anthropogenic drivers than the past. An observed trend in extremes could be because of a combination of some forcings (e.g., greenhouse gas emissions) and natural variability. In principle, one could add covariates such as El Niño Southern Oscillation to the GEV along with the trend component. However, this would be challenging since different physically-based covariates (and often more than one) will be necessary for different pixels/regions. More research efforts are needed in future to address these issues in non-stationary extreme value analysis.

Acknowledgements The authors would like to thank Professor Balaji Rajagopalan for his thoughtful comments on an earlier draft of this paper. We also acknowledge the comments of Dr. Francesco Serinaldi and two other anonymous reviewers which led to substantial improvements in the current version. This study is supported by the National Science Foundation (NSF) Award No. EAR-1316536, and the United States Bureau of Reclamation (USBR) Award No. R11AP81451. The first author acknowledges partial financial support from the National Center for Atmospheric Research (NCAR) Graduate Student Visitor Program. NCAR is sponsored by the National Science Foundation.

References

AghaKouchak A, Nasrollahi N (2010) "Semi-parametric and parametric inference of extreme value models for rainfall data.". Water Resour Manag 24:1229–1249

AghaKouchak, A., D. Easterling, K. Hsu, S. Schubert, and S. Sorooshian (2013)"Extremes in a Changing Climate", Springer, Netherlands.

Alexander, L.V., et al. (2006)"Global observed changes in daily climate extremes of temperature and precipitation." Journal of Geophysical Research 111. D5.

Beguería S et al (2011) "Assessing trends in extreme precipitation events intensity and magnitude using nonstationary peaks-over-threshold analysis: a case study in northeast Spain from 1930 to 2006.". Int J Climatol 31:2102–2114

Beniston M et al (2007) "Future extreme events in European climate: an exploration of regional climate model projections.". Clim Chang 81:71–95

Cannon AJ (2011) "GEVcdn: An R package for nonstationary extreme value analysis by generalized extreme value conditional density estimation network. Comput Geosci 37:1532–1533

Coles, S., (2001)"An introduction to statistical modeling of extreme values" Springer, London.

Coles, S., E.A. Powell (1996)"Bayesian methods in extreme value modelling: a review and new developments." International Statistical Review: 119–136

Cooley D (2009) "Extreme value analysis and the study of climate change.". Climatic Change 97:77–83

Cooley, D. (2013)"Return periods and return levels under climate change", Extremes in a Changing Climate. Springer Netherlands.

Cooley D et al (2007) "Bayesian spatial modeling of extreme precipitation return levels". J Am Stat Assoc 102: 824–840

Davison AC, Smith RL (1990) Models for exceedances over high thresholds. J R Stat Soc 393–442

Diffenbaugh NS, Giorgi F (2012) Climate change hotspots in the CMIP5 global climate model ensemble. Clim Chang 114(3–4):813–822

Easterling DR et al (2000) Climate extremes: observations, modeling, and impacts. Science 289:2068–2074

El Adlouni S et al (2007) Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. Water Resour Res 43

Field, C.B., et al., (2012)Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change. Cambridge University Press

Gelman, A., K. Shirley. (2011)"Inference from simulations and monitoring convergence" Handbook of Markov Chain Monte Carlo: 163–174.

Gelman, A., et al. (2003) "Bayesian data analysis" CRC Press

Gilleland, E., Katz, R.W. (2011)"New software to analyze how extremes change over time" Eos, 92 (2), 13—14.

Gumbel, E., (1942)"On the frequency distribution of extreme values in meteorological data". B. Am. Meteorol. Soc., 23.

Gumbel, E.J. (1958)"Statistics of Extremes", Mineola, NY: Dover.—, (1958)"Statistics of extremes". Columbia University Press, New York.

Hao Z et al (2013) "Changes in concurrent monthly precipitation and temperature extremes. Environ Res Lett 8(4):034–014

IPCC (2007)"Climate Change 2007: The Physical Science Basis", Working Group 1, IPCC Fourth Assessment Report, Cambridge University Press

Jakob D (2013) Nonstationarity in extremes and engineering design. Extremes in a Changing Climate, Netherlands

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90(430):773–795

Katz R (2010) Statistics of extremes in climate change. Clim Chang 100(1):71–76

Katz, R.W. (2013)"Statistical methods for nonstationary extremes". Extremes in a Changing Climate, Springer Netherlands.

Katz, R., et al., (2002)"Statistics of extremes in hydrology". Advances in Water Resources, 25, 12871304.

Kendall, M.G. (1976)"Rank Correlation Methods". 4thEd. Griffin.

Kharin VV, Zwiers FW (2005) "Estimating extremes in transient climate change simulations.". J Clim 18:1156–1173

Kharin VV et al (2007) "Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations.". J Clim 20:1419–1444

Klein T., et al. (2009) Guidelines on Analysis of extremes in a changing climate in support of informed decisions for adaptation. WMO-TD 1500, 56 pp. Leadbetter, M., et al. (1983)"Extremes and related properties of random sequences and processes".

Kundzewicz ZW, Robson AJ (2004) "Change detection in hydrological records—a review of the methodology". Hydrol Sci J 49(1):7–19

Li Y, Cai W, Campbell EP (2005) Statistical modeling of extreme rainfall in southwest Western Australia. J Clim 18(6)

Mann HB (1945) "Nonparametric tests against trend". Econometrica 13:245–259

Meehl GA et al (2000) "An introduction to trends in extreme weather and climate events: observations, socioeconomic impacts, terrestrial ecological impacts, and model projections". Bull Am Meteorol Soc 81: 413–416

Milly PCD et al (2008) Stationarity is dead: whither water management? Science 319:573–574

New M et al (2000) "Representing twentieth-century space-time climate variability. Part II: development of 1901–96 monthly grids of terrestrial surface climate.". J Clim 13:2217–2238

Papalexiou, S.M., D. Koutsoyiannis (2013)"Battle of extreme value distributions: A global survey on extreme daily rainfall." Water Resources Research.

Parey S et al (2010) "Different ways to compute temperature return levels in the climate change context.". Environmetrics 21:698–718

Renard, B., et al. (2006)"An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima." Water resources research 42.

Renard, B., et al. (2013)"Bayesian methods for non-stationary extreme value analysis", Extremes in a Changing Climate, Springer.

Rootzén H, Katz RW (2013) "Design life level: quantifying risk in a changing climate". Water Resour Res 49: 5964–5972

Rosbjerg, R. and Madsen, H. (1998)"Design with uncertain design values, Hydrology in a Changing Environment", Wiley, 155–163.

Salas JD, Obeysekera J (2013) Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. J Hydrol Eng. doi:10.1061/(ASCE)HE.1943-5584.0000820

Smith, R.L. (1987)"Estimating tails of probability distributions." Ann. Stat.: 1174–1207

Smith R (2001) "Extreme value statistics in meteorology and environment. Environmental statistics". Chapter 8: 300–357

Smith RL (1989) "Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone.". Stat Sci 4:367–377

Stephenson A, Tawn J (2004) "Bayesian inference for extremes: accounting for the three extremal types.". Extremes 7:291–307

Ter Braak CJF (2006) "A Markov chain monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces.". Stat Comput 16:239–249

Ter Braak CJF, Vrugt JA (2008) "Differential evolution Markov chain with snooker updater and fewer chains.". Stat Comput 18:435–446

Towler, E., B. Rajagopalan, et al., (2010)"Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory", Water Resour. Res., 46.

Villarini G et al (2009) "Flood frequency analysis for nonstationary annual peak records in an urban drainage basin". Adv Water Resour 32(8):1255–1266

Villarini, G., et al. (2011)"Annual maximum and peaks-over-threshold analyses of daily rainfall accumulations for Austria." Journal of Geophysical Research 116. D5.

Vose, R.S., D.R. Easterling, B. Gleason. (2005)"Maximum and minimum temperature trends for the globe: An update through 2004."Geophysical Research Letters 32.

Vrugt JA et al (2009) “Accelerating Markov chain monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling”. International Journal of Nonlinear Sciences and Numerical Simulation 10:273–290

Winkler, R.L. (1973)“A Bayesian approach to nonstationary processes”. IIASA.

Wu Z et al (2007) “On the trend, detrending, and variability of nonlinear and nonstationary time series.”. Proc Natl Acad Sci 104(38):14889–14894

Zhang X et al (2001) “Spatial and temporal characteristics of heavy precipitation events over Canada”. J Clim 14(9):1923–1936

Zhang X, Zwiers FW, Li G (2004) “Monte Carlo experiments on the detection of trends in extreme values.”. J Clim 17:1945–1952